Priors and prejudices

Priors and Prejudices: Comments on Susanna Siegel's *The Rationality of Perception* *

Andy Clark (University of Edinburgh, UK, and Macquarie University, Australia)

andy.clark@ed.ac.uk

Abstract

The Rationality of Perception (RP) depicts a kind of 'double counting' in which prior beliefs help to cause percepts that are then taken as evidence for those very beliefs. The correct response, RP argues, is to epistemically downgrade the percepts themselves, thus de-fanging their apparent role as rational support. I question RP's depiction of double-counting on the grounds that it fails to distinguish between cases involving the rational and the irrational use of prior information, and may lead to a kind of runaway epistemic downgrading. More positively, I suggest that to decide whether we confront a rational or irrational use of prior information, it helps to extend the temporal window and to bring world-altering action into the equation.

1. Introduction

The Rationality of Perception (henceforth, RP) is a stunning, elegant, astoundingly closely argued treatment. It raises new (and timely) issues, and makes a striking proposal. In this short treatment, I aim to highlight and engage just one aspect of this densely woven plot. That aspect is neatly summed up in the opening pages of RP with the question:

> "…if your prior beliefs could influence your experiences, *how could those very experiences go on to strengthen your beliefs*?"
> RP xiv (my emphasis)

To illustrate the puzzle, RP offers multiple possible examples of a kind of malicious influence in which prior beliefs seem to cause percepts that are wrongly taken as evidence for those very beliefs. One such case is loosely based on Dutch microscopist Antonie van Leeuwenhoek, who in 1677 first saw spermatozoa under the microscope. Already prone to preformationism (the idea that adult human bodies are fully but minutely preformed in human sperm), van Leewenhoek 'found' as visible in semen:

> "all manner of great and small vessels, so various and so numerous that I do not doubt that they be nerves, arteries and veins...And when I saw them, I felt convinced that, in no full grown body, are there any vessels which may not be found likewise in semen." From a letter quoted in Friedman 2002 p.76-77

A cleaner, but purely fictional, example is that of Jack and Jill (RP p.6, and throughout…):

> "Before seeing Jack, Jill fears that Jack is angry at her. When she sees him, her fear causes her to perceive Jack as angry, and this perception strengthens her fear."

Jill now feels that she has *more reason* for her fear. But is it rational for her to do so? Is it rational for her to believe her eyes and thus strengthen her prior (percept-inducing) belief that he is angry? On the one hand, it looks like Jill and the preformationist are just 'doing [their] best with the evidence' (RP p.6). But it also looks suspiciously like double-counting, as the 'Jack-is-angry' fear and the prior beliefs of the preformationists played some role in causing the experiences (of seeing visual signs of anger, of seeing tiny nerves and veins) that now seem to further justify those very beliefs (that Jack is angry, that sperm house tiny humans).

This is what Siegel calls the 'problem of hijacked experience' and results in the actual (raw) perceptual inputs not being given their "proper weight" (RP p.11). In such cases, the correct response, it is argued, is to epistemically downgrade the percepts themselves, thus de-fanging their apparent role as rational support. This means that we must now hold our own perceptual experiences to account. Instead of simply providing evidence that may freely be used to strengthen a belief, our perceptual experiences turn out to be as apt for epistemological assessment and critique as the judgments we base on them. Like judgments, they may be well- or ill-founded. Hence the titular conclusion, that perception itself is rationally assessable. Our perceptual experiences, if this is right, are not epistemically neutral after all. Instead, percept formation can itself be rational or irrational, as can the resulting experiences (see e.g. the thesis statement on RP p15). If RP is correct, then the epistemic status of a perceptual experience can be weakened by how it is formed or grounded (just like that of beliefs).

For the purposes of this essay, I'm going to simply assume that such effects (of belief, fear, etc. on bedrock perceptual experience) occur, and ask how we might think about them when they do. I think we have compelling

cognitive scientific reasons (Clark 2013; 2016) to think such effects are real – but these are not my focus. Instead, assuming they are real, I'll argue that the appearance of circularity and double-counting is – very often but not always- not to be feared, and ought not automatically result in epistemic downgrade.

In what follows I question RP's depiction of double-counting, mostly on the grounds that it fails to distinguish between cases involving the rational and the irrational use of prior information. More positively, I suggest that to decide whether we confront a rational or irrational use of prior information, it helps both to extend the temporal window and to bring world-altering action into the equation. In other words, we shouldn't just ask these questions about 'snapshot' perceptual moments but instead look at how the resulting percepts lead to exploratory actions that may (or may not) seem to confirm the percepts. It is this 'testing in action' that enables them quite properly, though never infallibly, to increase our confidence in the prior belief. If I'm right, then the initial appearance of circularity/double-counting is often a red herring, and by stressing it we may be exposing ourselves to a kind of runaway epistemic downgrading. In what follows, I try to steer a subtly different course between rationally active priors and (mere) prejudices.

2. Perception as Inference: Two Ways To Get Things Wrong

A good starting point is the increasingly accepted picture of perception as involving a process of sub-personal inference – a process in which prior knowledge meets sensory evidence, to generate a percept (Helmholtz 1860, MacKay 1956, Neisser 1967, Gregory 1980, Yuille and Kersten 2006, Friston et al 2012). This is a process that can deliver false results even when the processing itself is (arguably) 'optimal' in the way it combines priors and sensory evidence. I begin by looking at just such a case, before turning to some more puzzling ones that involve radically disturbed forms of processing and are more akin, I'll argue, to the epistemically unacceptable kinds of double-counting that RP rightly fears.

In a famous treatment, Weiss et al. (2002) show that a wide variety of motion illusions flow directly from the use of ideally rational ('Bayes-optimal') means of integrating priors and incoming sensory information. The illusions follow given only the ecologically reasonable priors that there is noise in the initial measurements, and that slower motions are more likely than faster ones. An intuitive example is driving in the fog. Here, drivers show a regrettable tendency to speed up and to underestimate their own and others velocities. This is because fog signifies low contrast (poor quality)

visual information about speed. Priors are biased toward slow speeds, as these are ecologically the norm. In the noisy sensory situation, these priors dominate and we may underestimate the true speeds of the cars. Weiss et al. (2002) conclude that:

> "Many motion "illusions" are not the result of sloppy computation by various components in the visual system, but rather a result of a coherent computational strategy that is optimal under reasonable assumptions". 2002, 603

These illusions result inevitably from the operation of a paradigmatically rational sub-personal regime that combines priors and (raw) sensory evidence in ways that work well in most cases. It is the intuitive *reasonableness* of this underlying (sub-personal) processing, not its doubtless debatable 'optimality', that matters for my argument. Even at the best of times, this is clearly a delicate balancing act. It is often reasonable to let prior knowledge cause you to ignore some incoming 'evidence' or apparent 'measurement' (treating it as noise) and instead to fill-in using expected values. If you don't do this, if you under-weight your priors, you will fail to spot faint patterns (like that famous Dalmatian against the spotty backdrop) hidden in noisy or ambiguous settings. But if you over-weight the priors, you will start to hallucinate patterns that are not there, just because you strongly expect them.

Just such an effect is shown in, for example, the so-called White Christmas experiments (Merckelbach and van de Ven 2001). Here, subjects were told that an audio clip would contain, hidden somewhere in the short segment, the rather faint onset of Bing Crosby's 'White Christmas'. The experimenters found that almost one-third of the subjects reported detecting the onset of the song. But the sound file contained only white noise – there was no trace of White Christmas hidden anywhere in the noise. The ability of some folk to 'detect' the familiar song is just an expression of an ability central to perceptual search and perceptual awareness in general: the ability to discount some aspects of a signal, treating them as 'noise', while accentuating other aspects, thus treating them as 'signal'. This ability, deployed under the influence of the strong expectation of a weak 'hard-to-detect' fragment of the familiar song, allows many perfectly normal subjects to enjoy what is in effect an auditory hallucination.

Neurally plausible process models such as 'predictive processing' (Bastos et al. 2012) depict these delicate balances between prior knowledge and sensory evidence as varying according to systemic estimates of the relative value or reliability ('precision') of each source – the priors, and the various

bits of prima facie sensory evidence) in context. This involves the use of neuromodulators such as dopamine to vary the post-synaptic weighting, or gain, on the incoming sensory information versus the prior-based 'top-down' prediction.

This is a long story, with many moving parts. But one key, relevant, idea (Griffin and Fletcher 2017) is that many forms of psychosis and aberrant experience may indeed involve a vicious cycle – very much akin to that which RP depicts as 'hijacked experience' – in which priors and sensory experience become locked in a self-entrenching spiral whereby subtly altered high-level beliefs help construct subtly altered percepts that appear to confirm and reinforce those very beliefs.

This kind of spiral plausibly occurs in some cases of schizophrenia, where early false inferences lead to subtly altered percepts that in turn seem to evidence and entrench increasingly bizarre beliefs. In this way, the two 'positive symptoms' of schizophrenia (delusions and hallucinations) emerge as co-emerging and co-determining aspects of a single fault in a multi-level (Bayesian) inference regime. The improbable (replacement, telepathy, conspiracy) slowly but surely becomes the least surprising, as false perceptions and bizarre beliefs become locked into an epistemically insulated, self-confirming cycle (Fletcher and Frith 2009).

Griffin and Fletcher (2017) describes how the same kind of process could explain some cases of Capgras delusion – the belief that your loved one has been replaced by an imposter. This is a condition that may be seeded by a lack (however caused) of some sub-personally predicted physiological response in their presence. But this seed may lead, the authors argue, to a spiral of false inference that sounds uncannily reminiscent of Siegel's Jack and Jill thought experiment. It is worth quoting their description in some detail:

> "The first experience may consist of nothing more strange or specific than a mild sense of jamais vu that leads one to fleetingly imagine the impostor idea without taking it seriously. However, on the second occasion ….merely having entertained the impostor idea at a high level [increases its plausibility….. Perhaps, this time, the [loved one] seems not merely unfamiliar but also suspicious (e.g., their smile is now perceived perhaps as having a mocking or sinister quality)…In Bayesian terms, the impostor hypothesis is *more probable given this second experience than it was given the first experience*." 2017 p.279

Griffin and Fletcher cite multiple known effects that could play a role in this kind of process. 70% of Capgras cases occur in individuals diagnosed with

schizophrenia. Even for neurotypical agents, the "mere act of imagining a hypothesis raises its subjective probability next time it is entertained" (Arkes et al. 1991, Garry et al. 1996, Goff&Roediger 1998) even "when what is imagined is bizarre and a priori implausible" (Seamon et al. 2006, Thomas & Loftus 2002). The 'imagination inflation effect' is "stronger the more vividly the hypothesis was imagined" (Sherman et al. 1985, Thomas et al 2003), and schizophrenic subjects display a strong tendency towards vivid mental imagery (Oertel et al. 2009). Finally, the effect is greater for emotional interpersonal events than for neutral ones (Szpunar & Schacter 2013), which speaks to the emotionally charged nature of many delusions.

In other words, multiple effects that occur even in neurotypical processing are (in the case of psychosis) being co-opted as part of a rolling process in which inference goes steadily astray. We do not confront the sudden emergence of a full-blown psychosis. Instead, when neuromodulatory systems are radically disturbed, what is compromised is the temporally extended process of perceptual inference itself. Now, looping effects of the kind feared by Siegel do indeed emerge. Describing these, Jardi and Deneve (2013) suggest that:

> "bottom-up sensory information and top-down predictions are *reverberated*: prior beliefs are misinterpreted as sensory observations and vice versa. As a result, these predictions are counted multiple times" 2013, 3227

Jardri and Deneve call this 'circular belief propagation' (see also Jardi et al 2017).

All this sets the scene for the question I wish to pursue. In RP style cases of 'perceptual hi-jacking' do we see at work something more like the Weiss scenario, where the combination of priors and sensory evidence leads to generally correct outcomes in the ecologically normal environment? Or is it more like the psychosis model, in which disturbed neuromodulatory economies yield disturbed patterns of inference over time, allowing priors and percepts to become progressively more and more tightly locked into a reverberating, mutually misleading dance?

3. Bedrock City – Some Thought Experiments

I think we can make progress on this by imagining a couple of scenarios involving ordinary perception.

Scenario 1: Seeing Dino

You look out the window of your house in Bedrock City and see a familiar-looking shape. Your high-level belief that the object most likely to be in the yard is your pet dinosaur, Dino, then programs a very specific set of visual saccades that aim to harvest evidence for that very hypothesis. Your exploratory saccades aim to discover a specific shape, and pay special attention to that distinctive perky tail. If the object is indeed Dino, you should get that very evidence, allowing the system to settle into a stable percept. It seems entirely right that at that point your systemic faith in the Dino hypothesis increases. This is exactly the (paradigmatically rational) process described in Friston et al 2012.

Scenario 2: Not Seeing Dino

Next, let's assume that it is a foggy night, or you are a little distracted, or didn't get quite enough sleep. As before, your strong prior belief that it is Dino in the yard programs a sequence of exploratory saccades. You briefly explore the scene, the test is passed, you seem to see the outline of Dino, perky tail and all. Systemic faith that it is Dino increases. But this time you are wrong. In fact, it is Baby Puss, your sabre-toothed cat again, trying to sneak back into the house.

How should RP treat the situation here? Should we, under such circumstances, not allow the resulting (false) Dino percept, that likewise passes the test of programming saccades that seem to confirm the hypothesis, to properly increase your degree of belief that it is Dino in the yard? Should we impose an epistemic downgrade on that percept, blocking it from rationally increasing your confidence in the belief that Dino is in the yard? It is true that, just as in the case of angry Jack, the percept was partly caused by the belief that it then cements. But this was also true in the previous, and (I suggest) paradigmatically rational scenario.

I suggest that downgrade can be justified in neither of the two cases. Epistemically, the cases are on a par - each depicts a reasonable response to the evidence, given the priors, and each confronts (and passes) the tribunal of ongoing testing by programming actions.

To press this, consider the case where, in Scenario 2, the animal (wrongly identified as Dino) suddenly starts to roar or meow, and I rapidly re-parse the foggy scene, seeing the sabre-toothed cat. The route to this rationally revised percept still strongly implicates my priors - that pesky cat was, after all, the next-most-probable cause. But I am surely right to take the success of the cat hypothesis at accommodating these new waves of sensory evidence as itself increasing the subjective probability that the object is my cat. To think otherwise would be to deprive me of a standard, and crucial,

route to rational belief revision. The upshot is that we cannot downgrade the second case without downgrading the (paradigmatically rational) revision-variant too.

The bigger picture here involves appreciating that percepts are simply not ends-in-themselves. They are tools that lead us to sample the world in ways that aim to minimize long-term uncertainty. As such, they are tested in on-going actions (including visual saccades) and as long as they pass the tests the guiding hypothesis gains in plausibility. Their epistemic value is thus defined not by their snapshot content but by their role in guiding temporally extended cycles of perception and action. What matters here, as in science itself, is the longer-term openness of the system (via action) to harvesting and recognizing disconfirming evidence. Just because what we expect to see influences what we do see, that doesn't block that possibility. And as long as it doesn't, the influence of prior belief on prior-belief confirming percept seems benign.

In closing, note that even if you buy all this, you can still ask about the etiology of the priors themselves. This might provide at lest some of the leverage that RP rightly seeks regarding cases such as those discussed in Chapter 10 concerning the mistaken perception of a cellphone as a gun when held by a black man in a seedy alleyway. What may be at issue between us here is the allocation of blame. RP, in such cases, seems to be aiming epistemic reproach not at on-board information processing regimes nor solely at the cultural milieu but also but at the individuals themselves. For it is the individuals who are held to be reasonable or unreasonable. However, I can't see any way to judge the individuals apart from as culturally situated cognitive engines exhibiting (or failing to exhibit) certain checks and balances in the use of sensory evidence and priors.

Siegel argues (in chapter 10, using the useful image of the 'mind of the world') that perceptual epistemic downgrade can be simply inherited from a bad sample or a statistically unrepresentative environment. I'm strongly inclined to accept this. To me, this locates primary blame right where it belongs, in the distortive cultural milieu, but lets a form of epistemic damage (though not epistemic culpability) be inherited in the form of those badly-grounded priors. Perhaps we can even go further (as Larissa MacFarquhar has suggested to me in conversation) and treat some disturbances to the 'mind of the world' as analogous to flawed reverberating inference in schizophrenia. Consider, for example, the 'fake news meme'. This systematically upsets the delicate balances between priors and evidence, making room for catastrophic reverberating societal influences that rather closely resemble those found in individual psychosis. For a powerful and

epistemologically important analysis of such effects, see Vosoughi et al 2018.

4. Conclusions: Of Bayes and Bathwater

Siegel might (or might not) agree. By the time we get to RP 138-9, the story is that inference-based, prior-reflecting influence on percept-formation is legitimate iff the pre-conscious perceptual input is 'given proper weight' in the processing. But exactly what does this imply?

RP claims that Jill's 'hijacked experience' loses power as support of the belief that Jack is angry, because of the role of her jack-is-angry prior in making the percept one of a subtly angry looking Jack. (see p.21). But this, I have argued, is really just Bayesian business as usual, allowing prior beliefs to tip an otherwise delicate balance**.** It doesn't seem to me that the ongoing processing in the individual is in any way at fault here. Some evidence for this less damaging diagnosis is provided by thinking about temporally extended processing cycles that might (or would not) respond to new evidence in apt ways. If instead we follow RP, and epistemically downgrade the Angry Jack case, then surely we ought to downgrade the Sabre-tooth cat case (section 3) too. But to do so leads, I claim, to a real threat of runaway epistemic downgrading – of throwing out rational, hypothesis-testing Bayes with the unwanted bathwater.

Fortunately, there is a clear and promising alternative. The alternative is to fault the individual inference engine only when it is systematically reverberating very faint evidence or very strong priors, becoming progressively more and more closed to new information. We would then worry about the appearance of double-counting only in cases that have more in common with runaway psychosis than with the (always ampliative and hence always fallible) use of prior knowledge to speed processing, reduce ambiguity, and separate newsworthy signal from noise.

* This paper was written with support from ERC Advanced Grant XSPECT - DLV-692739 ('Expecting Ourselves: Prediction, Action, and the Embodied Mind').

References

Hal Arkes, Stephan Boehm LE, and Gang Xu 1991. "Determinants of judged validity". *J. Exp. Soc. Psycho*l. 27(6):576–605

Andre Bastos, W. Martin Usrey, Rick Adams, George Mangun, Pascal Fries, and Karl Friston. 2012 "Canonical microcircuits for predictive coding" *Neuron* 76:695-711.

Andy Clark 2013. "Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science" *Behavioral and Brain Sciences* 36: 3: p. 181-204

Andy Clark. 2016. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind* New York, Oxford University Press

Marc Ernst and Martin S. Banks. 2002 "Humans Integrate Visual and Haptic Information in a Statistically Optimal Fashion. *Nature* 415: 429-433

Paul Fletcher and Chris Frith. 2009 "Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia". *Nature Reviews: Neuroscience* 10: 48-58

David M. Friedman. 2002 *A Mind of Its Own: A Cultural History of the Penis*. New York: The Free Press

Karl Friston, Rick Adam, Laurent Perrinet, and Michael Breakspear. 2012 "Perceptions as hypotheses: saccades as experiments". *Front. Psychology* 3:151.

Maryanne Garry, Charles Manning, Elizabeth Loftus, and Jim Sherman. 1996. "Imagination inflation: imagining a childhood event inflates confidence that it occurred." *Psychon. Bull. Rev.* 3(2):208–14

Lynn Goff and Henry Roediger. 1998. "Imagination inflation for action events: Repeated imaginings lead to illusory recollections". *Mem. Cogn.* 26(1):20–33

Richard Gregory 1980. "Perceptions as hypotheses". *Phil. Trans. R. Soc. Lond., Series B, Biological Sciences* 290(1038): 181-197.

Juliet Griffin & Paul Fletcher 2017. "Predictive Processing, Source Monitoring, and Psychosis". *Annual Review of Clinical Psychology*, *13*(1), 265–

289.

Hermann von Helmholtz. 1860/1962 *Handbuch der physiologischen optik, vol. 3*, ed. &trans. J. P. C. Southall. Dover. (Original work published in 1860; Dover English edition in 1962).

David MacKay. 1956. "The epistemological problem for automata". CE Shannon & J McCarthy, editors, *Automata Studies,* Princeton, NJ: Princeton University Press, 235-251.

Harold Merckelbach & Vincent van de Ven 2001. "Another White Christmas: fantasy proneness and reports of 'hallucinatory experiences' in undergraduate students". *Journal of Behaviour Therapy and Experimental Psychiatry*, 32, 137-144.

Ulrich Neisser. 1967. *Cognitive Psychology*. New York, Appleton-Century-Crofts

Viola Oertel, Anna Rotarska-Jagiela,Vincent van de Ven, Corinna Haenschel Michael Grube, Ulrich Stangier, Konrad Maurer, and David E.J. Linden. 2007 "Mental imagery vividness as a trait marker across the schizophrenia spectrum" *Psychiatry Res*. 167(1):1–11

John G. Seamon, Morgan M. Philbin, and Liza G. 2006. "Do you remember proposing marriage to the Pepsi machine? False recollections from a campus walk". *Psychon. Bull. Rev*. 13(5):752–56

Steven J. Sherman, Robert B. Cialdini, Donna F. Schwartzman, and Kim D. Reynolds. 1985. "Imagining can heighten or lower the perceived likelihood of contracting a disease: the mediating effect of ease of imagery". *Pers. Soc. Psychol. Bull.* 11:118–27

Karl Szpunar and Dan Schacter. 2013. "Get real: effects of repeated simulation and emotion on the perceived plausibility of future experiences". *J. Exp. Psychol. Gen*. 142:323–27

Ayanna Thomas and Elizabeth Loftus. 2002. "Creating bizarre false memories through imagination." *Mem. Cogn*. 30(3):423–31

Ayanna Thomas, John Bulevich, and Elizabeth Loftus E. 2003. "Exploring the role of repetition and sensory elaboration in the imagination effect".

Mem. Cogn. 31:630–40

Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018 "The spread of true and false news online." *Science*, *359*(6380), 1146 LP-1151.

Yair Weiss, Eero P. Simoncelli & Edward H. Adelson. 2002 Motion Illusions as Optimal Percepts *Nature Neuroscience* 5: 6: 598 – 604

Alan Yuille and Daniel Kersten. 2006 "Vision as Bayesian inference: analysis by synthesis?" *Trends in Cognitive Science*: 10: 7: 301-308